

## **Data Mining Project** ***From Data Analysis to Knowledge Extraction***

This Data Mining project aims to help you become familiar with some typical tasks of the data analyst and data scientist profession. In addition to the analysis, cleaning, integration, selection and transformation of data, there are three main axes of data mining that allow the extraction of knowledge, namely: the extraction of frequent patterns, classification and clustering.

### **Information:**

- Work in pairs.
- The list of pairs must be sent on the [google form](#) shared in the google classroom.
- Report to be submitted in digital + paper version for each of the 2 parts.
- A presentation of the interface and the source code will be organized at the end of each part of the project.
- It is important to provide a well-structured and easy-to-read report.
- Important information and conclusions must be highlighted, provide an adequate analysis of the graphs presented, etc.
- Take advantage of the practical sessions (TP) to ask questions rather than by email.
- Any initiative can be rewarded!

### **Datasets:**

This project necessitates two (2) datasets that are available on the google classroom:

- **Soil-DATA** : which contains soil properties of the Algerian territory.
- **Climate-DATA** : which contains climate variables of 2019 for the entire world.
- **Wilayas** : which contains the shape of Algeria Wilayas (Polygon) to help you extract the Climate data of Tizi-ouzou only.